HUK-COBURG

Spectrum Scale HDFS-Transparency Connector UseCase: Backup of Hadoop Clusters at HUK-Coburg

HDFS-Spectrum Scale Details and Status

Coburg, 04.03.2021

Agenda:

- Introduction HUK-Coburg
- Review last year Presentation
- Implementation-Details
  - Transparency-Config
  - Scale-Config
  - Some issues
- Time-line/Solution Experiences
- What‘s Next

HUK-COBURG

The parent company HUK-COBURG a. G.

Mutual

insurance company

Operating in the insurance sector for

87

years

Only for

public servants

Principle of mutuality:

The company is owned by the insured persons.

The object of the company is solely geared to the interests of the insured persons.

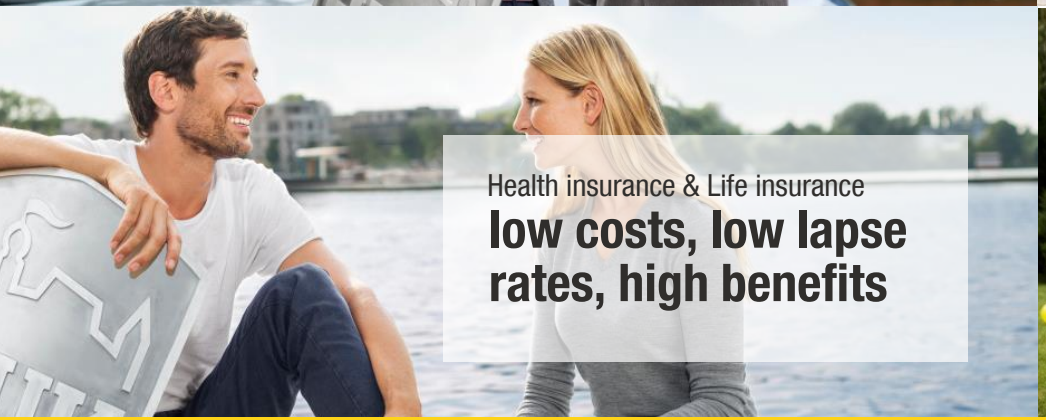
3.6

million members

The entire group works according to the

principle of mutuality

Introduction HUK-Coburg



Introduction HUK-Coburg

HUK-COBURG

Being close to the customer is of utmost importance to us ...

Multi-channel approach
free choice for our customers

In 38 branch offices, about 2,500 permanent employees
offer expert advice, underwriting and claims service.

Nearly 680 customer service offices:
self-employed full-time field service

Around 3,000 local part-time "trusted counsellors"
near our customers

More than 100 consultants
specialising in life, health and accident insurance

Customer service centres
1,500 customer service agents provide telephone support on policy matters.

Comprehensive range of insurance products available on the Internet
Our online-only subsidiary HUK24 offers particularly favourable prices.

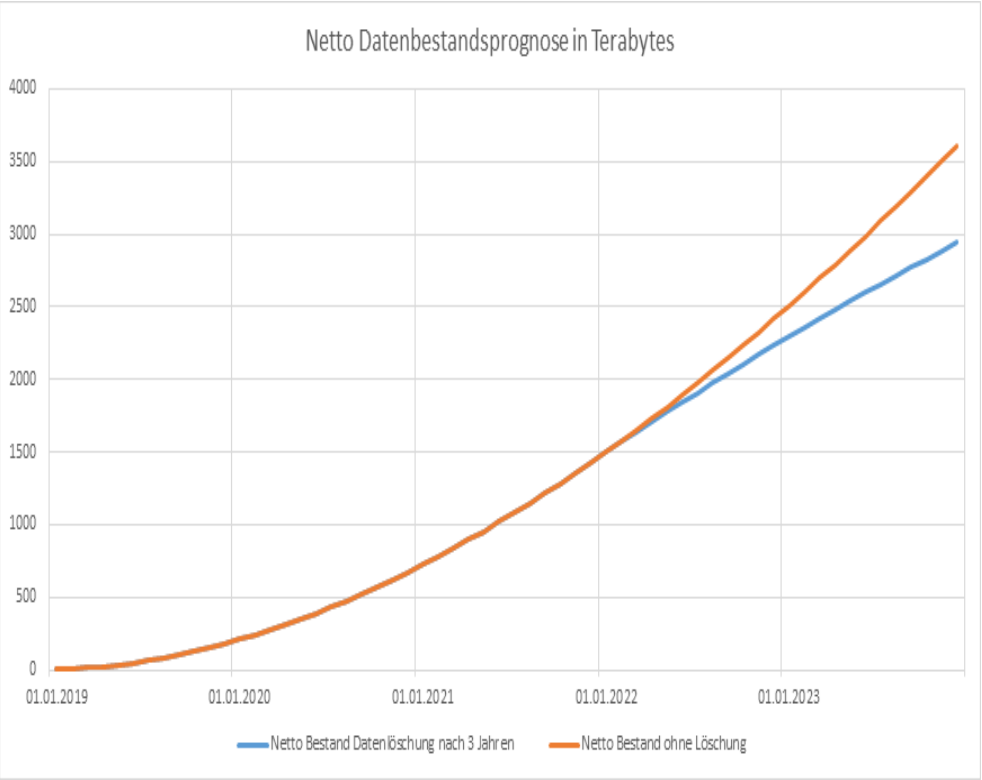
Versicherer im Raum der Kirchen (Insurers in the field of churches)
about 320 local customer contact persons for church-related staff

pag 5

Review last year presentation

- Backup-Interface to all hadoop-clusters
- Used at our telematik use case
- Tiering of hadoop data
- Implementation of cost effective storage media

- Backup-Requirments
  - daily backup of Hbase scoring
  - daily differential backup of TripData with media break
  - daily backup of analytics results



HUK-COBURG

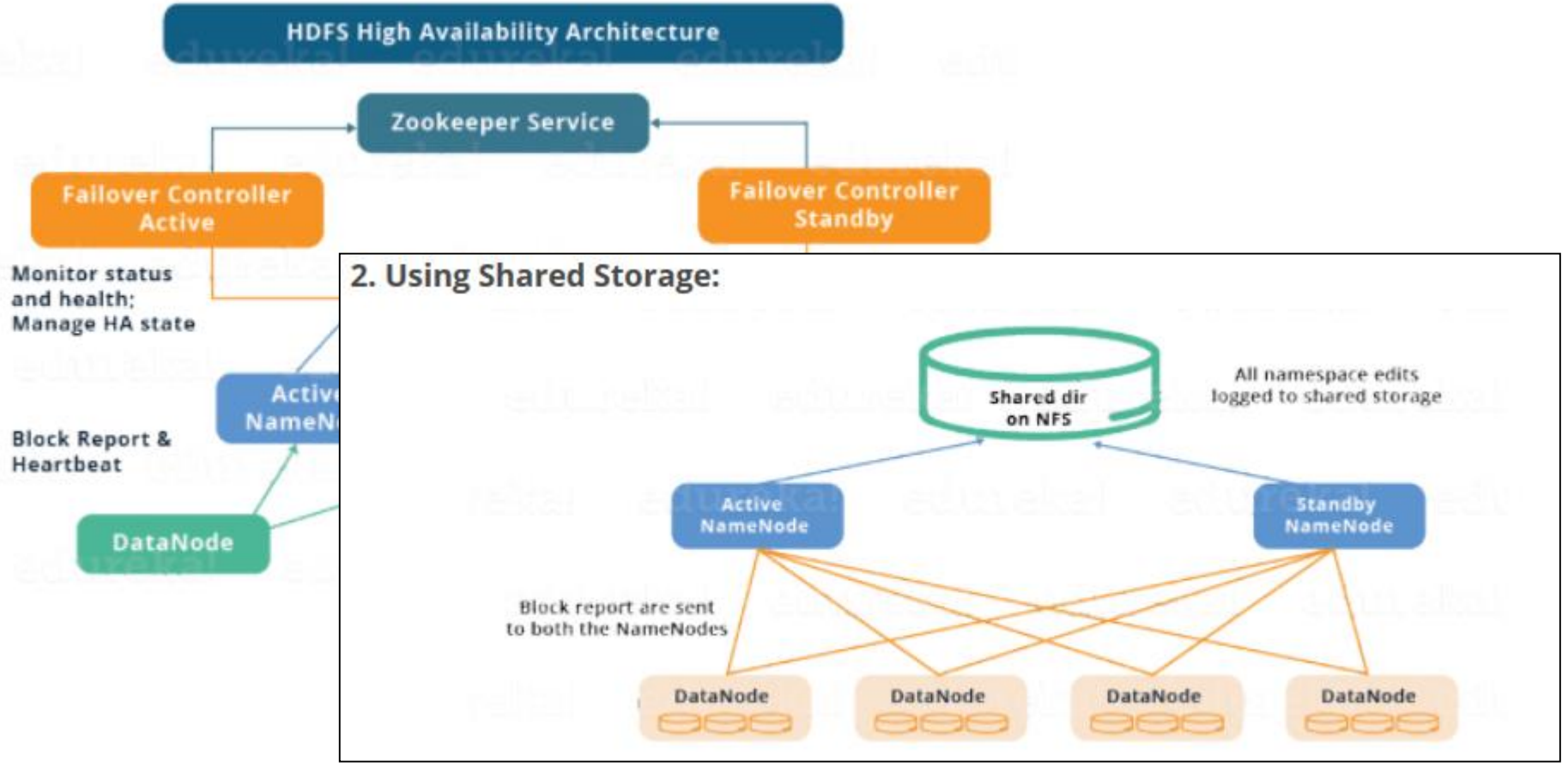
Hadoop Features in use

- Name-Node HA
  - zookeeper and ZKFController Config
  - Spnego/Kerberos Config
  - Dual Namenode Config
- Namenode Federation with usage of discp from source-clusters
  - distcp from native Cloudera/HDP-Cluster to HDFS-Transparency
- Ambari only for Handling Kerberos and Zookeeper
  - for regenerate Keytabs and Principal renewal
- Transparency Connector in Version gpfs.hdfs-protocol-3.1.0-5.x86_64

  Testsystems with gpfs.hdfs-protocol-3.1.0-7.x86_64

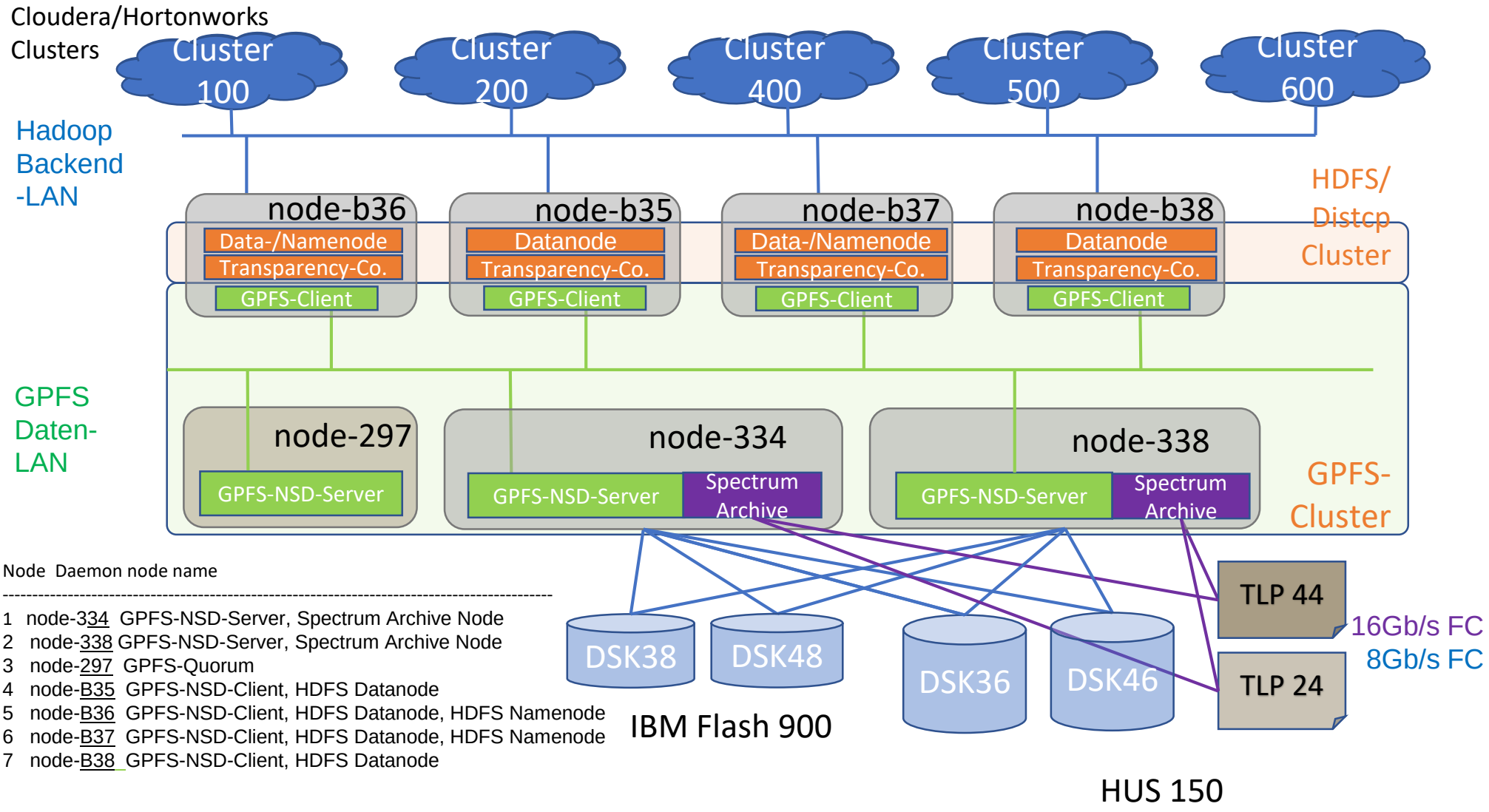
HUK-COBURG

Hadoop HA-Features Overview traditional vs. Scale transparency



HUK-COBURG

Architecture Overview- Overview



Node Daemon node name

---

1 node-334 GPFS-NSD-Server, Spectrum Archive Node
2 node-338 GPFS-NSD-Server, Spectrum Archive Node
3 node-297 GPFS-Quorum
4 node-B35 GPFS-NSD-Client, HDFS Datanode
5 node-B36 GPFS-NSD-Client, HDFS Datanode, HDFS Namenode
6 node-B37 GPFS-NSD-Client, HDFS Datanode, HDFS Namenode
7 node-B38 GPFS-NSD-Client, HDFS Datanode

HUK-COBURG

Konfiguration-Details:

- configration-path scale
  - /var/mmfs/hadoop/etc/hadoop/
    - core.-site.xml
    - hdfs-site.xml
    - gpfs-site.xml
- configuration path hadoop
  - /etc/hadoop/conf

- importend cmd‘s
  - /usr/lpp/mmfs/hadoop/bin/hdfs haadmin -getAllServiceState
  - /usr/lpp/mmfs/hadoop/bin/hdfs haadmin -failover nn1 nn2
  - mmhadoopctl connector [status, stop, start]
  - jps

HUK-COBURG

Some issues/ Hints

- Authentication/authorisation Points
  - in difference to native Hadoop (CDH,HDP) implementations in transparency all uid/guid must be the same on all transparency nodes
  - hadoop-keytabs should only the service-principals defined and no sharing with system keytabs if possible
  - check your kerberos-credential policies (lifetime of entries)
  - hadoop.security.auth_to_local function for principal identication
- Update errors
  - we had trouble on 3.1.0.6 because of point 1 above. (plan with 3.1.0.7)
  - Generate a Test-Environment with same Version to go forward and evaluate upgrades.

HUK-COBURG

- Production since Aug 2019
  - only 1 issues after version upgrades from 3.1.0.4 to 3.1.0.6
  - currently 2,2 PB managed with Scale-Transparency and Spectrum Archive at Backend
  - Tiered storage managed with DMAPI (HSM)

Capacity View:

Whats Next?

Migration Planing to go to CDP because of EOS HDP

- *requirements are scale >=5.1.0.1 and hdfs-CES*
- ***Missing**: Migration-plan from HDP Transparency to CES-HDFS CDP.*
- *rhel7.9 Support ist currently for HDP missing.*

Migration to our ESS-Backup-Environment

- Migration Storage to ESS

Test Automation integration of HDFS-CES-Service

- *ansible playbooks for steps to integrate ces-hdfs in Cloudera Manager*

Containerisation of Protocollservice

- *Are hdfs also in plan to implement these on CNSA??*

Questions ?